JOURNAL OF FUNCTIONAL ANALYSIS 77, 198–210 (1988)

A Szegö Limit Formula for Conjugates of ψDO's by Hecke Operators

STEVEN ZELDITCH

Department of Mathematics, University of California, Berkeley, California 94720

Communicated by the Editors

Received July 8, 1986

Let T_g be a Hecke operator and let $\mathrm{Op}(a)$ be a ψDO on $L^2(\Gamma\backslash\mathfrak{h})$. Further, let π_λ be the projection onto span $\{u_j\colon -\Delta u_j=\lambda_j u_j,\ T_g u_j=\rho_j(g)\,u_j,\ \sqrt{\lambda_j}\leqslant\lambda\}$. We will determine the Szegö limit measure $\mu_{g,a}$ for the operators $T_g^*\,\mathrm{Op}(a)\,T_g\colon \int_{-\infty}^{\infty} x^k d\mu_{g,a}(x)=\lim_{\lambda\to\infty}(1/\mathrm{rank}\,\pi_\lambda)\ \ \mathrm{tr}(\pi_\lambda T_g^*\,\mathrm{Op}(a)\,T_g\pi_\lambda)^k$. Comparison of $\mu_{g,a}$ with the Szegö limit measure of $\mathrm{Op}(T_g a)$ shows that $T_g^*\,\mathrm{Op}(a)\,T_g$ is quite remote spectrally from $\mathrm{Op}(T_g a)$. © 1988 Academic Press, Inc.

Let X be a finite area hyperbolic surface $\Gamma\backslash\mathfrak{h}$ ($\mathfrak{h}$ = upper $\frac{1}{2}$ plane, $\Gamma\subset PSL_2(\mathbb{R})$). Suppose there is a $g\in PSL_2(\mathbb{R})$, $g\notin\Gamma$, so that $\Gamma'(g)=g^{-1}\Gamma g\cap\Gamma$ has finite index in both Γ and $g^{-1}\Gamma g$, and that $g\notin\Gamma$. Then one has a Hecke operator $T_g\colon L^2(X)\to L^2(X)$; namely if Γ is the disjoint union $\bigcup_{j=1}^{d}\Gamma'(g)\,\gamma_j$, $T_g f(x)=\sum_{j=1}^{d} f(g\gamma_j x)$ (note: the sum is frequently normalized). Geometrically, a Hecke operator T_g is a Radon transform arising from a diagram of finite locally isometric covering projections:

$$\begin{tikzcd} & \Gamma'(g)\backslash\mathfrak{h} \arrow[dl, "\pi"'] \arrow[dr, "\rho"] & \\ \Gamma\backslash\mathfrak{h} \arrow[rr, leftrightarrow, "l_g"] & & g^{-1}\Gamma g\backslash\mathfrak{h} \end{tikzcd}$$

(l_g is the canonical identification of $\Gamma\backslash\mathfrak{h}$ and $g^{-1}\Gamma g\backslash\mathfrak{h}$ induced by left translation by g on $\mathfrak{h}$). Thus $T_g=l_g\rho_*\pi^*$.

More generally, T_g extends to an operator on $C^\infty(S_*X)=C^\infty(\Gamma\backslash PSL_2(\mathbb{R}))$ via the diagram:

0022-1236/88 $3.00

$$\begin{tikzcd}[column sep=small] & \Gamma'(g)\backslash PSL_2(\mathbb{R}) \arrow[dl, "\pi"'] \arrow[dr] & \\ \Gamma\backslash PSL_2(\mathbb{R}) \arrow[rr, leftrightarrow, "l_g"] & & g^{-1}\Gamma g\backslash PSL_2(\mathbb{R}). \end{tikzcd}$$

T_g may then be extended by homogeneity to $C^\infty(T^*X)$.

Now let Δ be the Laplacian on X. Then $L^2(X)=L^2_{\rm d}(X)\oplus L^2_{\rm c}(X)$, where $L^2_{\rm d}$ is the discrete spectral subspace and $L^2_{\rm c}$ the continuous spectral subspace. Clearly $[\Delta, T_g]=0$. We will assume henceforth that $[T_g, T_g^*]=0$ also (this is satisfied by all T_g coming from arithmetic Γ). Then $L^2_{\rm d}(X)$ has an orthonormal basis of simultaneous eigenfunctions $\{u_j\}$ for T_g and Δ,

$$\Delta u_j=-\lambda_j^2 u_j,$$
$$T_g u_j=\rho_j(g)\, u_j.$$

Recently, P. Sarnak has determined the Szegö limit distribution of the eigenvalues $\rho_j(g)$ for arithmetic Γ. By definition, this distribution is the measure μ_g on $C(\mathbb{R})$ satisfying:

$$\operatorname{tr}_\Delta f(T_g)=\int_{\mathbb{R}} f(x)\, d\mu_g(x),$$

where $\operatorname{tr}_\Delta f(T_g)=\lim_{\lambda\to\infty}(1/N(\lambda))\sum_{\lambda_j\leqslant\lambda} f(\rho_j(g))$, and $N(\lambda)=\#\{\lambda_j\leqslant\lambda\}$. Otherwise, put $\operatorname{tr}_\Delta f(T_g)=\lim_{\lambda\to\infty}(1/N(\lambda))\operatorname{tr} f(\pi_\lambda T_g\pi_\lambda)$ where π_λ is orthoprojection onto $\operatorname{span}\{u_j\colon \lambda_j\leqslant\lambda\}$. Sarnak's theorem states that the moment $\int_{\mathbb{R}} x^m\, d\mu_g(x)=M(g,m)$ is equal to the number of words $\alpha_{i_1}\cdots\alpha_{i_m}\equiv I(\Gamma)$, where α_i are the coset representatives $g\gamma_i$. For many Γ and g, either $\operatorname{tr}_\Delta T_g^m$ or $M(g,m)$ can be computed explicitly and the inverse moment problem can be solved to find μ_g. For example, when $\Gamma=SL_2(\mathbb{Z})$, $g=\begin{pmatrix} p^{1/2} & 0\\ 0 & p^{-1/2}\end{pmatrix}$ and $T_p f(z)=(1/\sqrt{p}\sum_{ad=p,\, b \bmod p} f((az+b)/d)$, Sarnak proves that $d\mu_p$ is the "spherical Plancherel measure for $SL_2(\mathbb{Q}_p)$,"

$$d\mu_p(x)=\begin{cases} 0 & |x|>2\\ \dfrac{(1+p)\sqrt{4-x^2}}{2\pi(n(p)^2-x^2)}, & |x|<2 \quad (n(p)=p^{1/2}+p^{-1/2}).\end{cases}$$

Our purpose in this paper is to generalize Sarnak's theorem for the case of $SL_2(\mathbb{Z})$, although it is plausible that our results are true for all arithmetic $\Gamma\subset PSL_2(\mathbb{R})$. What we do is to consider $\operatorname{tr}_\Delta f(\pi_\lambda T_g \operatorname{Op}(a)\, T_g^*\, \pi_\lambda)$ where $\operatorname{Op}(a)$ is a zeroth order self-adjoint pseudodifferential operator (ψDO). Let $d\mu_{g,a}$ be the Szegö limit measure:

$$\operatorname{tr}_\Delta f(\pi_\lambda\, T_g \operatorname{Op}(a)\, T_g^*\, \pi_\lambda)=\int f(x)\, d\mu_{g,a}(x).$$

Our main result is a reasonably explicit formula for the moments $\int x^m \, d\mu_{g,a}$ when $\Gamma = SL_2(\mathbb{Z})$ and $g = \left(\begin{smallmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2} \end{smallmatrix}\right)$. To state the result, we need to set up some notation and background. First, $\Gamma = \bigcup_{j=1}^{d} \Gamma'(g)\gamma_j$ if and only if $\Gamma g \Gamma = \bigcup_{j=1}^{d} \Gamma \alpha_j$ where $\alpha_j = g\gamma_j$. Second, the α_j can be chosen to be two-sided: $\Gamma g \Gamma = \bigcup_{j=1}^{d} \Gamma \alpha_j = \bigcup_{j=1}^{d} \alpha_j \Gamma$. For the case of $SL_2(\mathbb{Z})$, we will fix such $\alpha_j = \left(\begin{smallmatrix} -jp^{-1/2} & -p^{1/2} - j^2 p^{-1/2} \\ p^{-1/2} & jp^{-1/2} \end{smallmatrix}\right)$, $0 \leqslant j \leqslant p-1$, $\alpha_p = \left(\begin{smallmatrix} 0 & p^{-1/2} \\ -p^{1/2} & 0 \end{smallmatrix}\right)$ for the $p+1$ cosets. We will need the following:

LEMMA. *Let $\{\alpha_j\}$ be the above coset representative for $\Gamma \left(\begin{smallmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2} \end{smallmatrix}\right) \Gamma$. Then a word $\alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m}$ is in $SL_2(\mathbb{Z})$ if and only if it reduces to the identity.*

Our main result is then

THEOREM. *Let $\Gamma = SL_2(\mathbb{Z})$ and $g = \left(\begin{smallmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2} \end{smallmatrix}\right)$. Then*

$$\lim_{\lambda \to \infty} \operatorname{tr}(\pi_\lambda T_g \operatorname{Op}(a) T_g^* \pi_\lambda)^m$$

$$= \int_F \sum_{\substack{(i_1, \dots, i_m) \\ (j_1, \dots, j_m) \\ \alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m} = e}} a_0(\alpha_{j_m} \xi)\, a_0(\alpha_{j_{m-1}} \alpha_{o_m}^{-1} \alpha_{j_m} \xi) \cdots a_0(\alpha_{j_1} \alpha_{i_2}^{-1} \cdots \alpha_{j_m} \xi)\, d\omega(\xi),$$

where F is a fundamental domain for Γ in $PSL_2(\mathbb{R})$, $d\omega$ is Haar measure, and $\alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m}$ fully reduces to e. Use of the above lemma then makes it straightforward to evaluate any given moment. For example,

(i) $\operatorname{tr}_\Delta(T_g \operatorname{Op}(a) T_g^*) = \int_F (T_g a)\, d\omega$,

(ii) $\operatorname{tr}_\Delta(T_g \operatorname{Op}(a) T_g^*)^2 = \int_F (d-1)(T_g(a^2) + (T_g a)^2)\, d\omega$.

It looks quite difficult to give a general formula for the mth moment in terms of T_g and a, and to solve the inverse moment problem for $d\mu_{a,g}$.

This problem arose as an attempt to understand the effect of conjugating ψDO's by Hecke operators. As will be seen in Section 2, one has an Egorov type theorem: $T_g \operatorname{Op}(a) T_g^* = \operatorname{Op}(T_g a) + R_g(a)$, where $R_g(a)$ is a certain Fourier Integral Operator. When $T_g = T_p$ for $SL_2(\mathbb{Z})$ and $a = 1$, the well-known formula $T_p^2 = I + (1/p)\, T_{p^2}$ shows that $R_p(1) = (1/p)\, T_{p^2}$. As a first step in understanding the spectral properties of the general $R_p(a)$ we considered its Szegö limit distribution, which will be deduced, like that of $T_p \operatorname{Op}(a) T_p^*$, in Section 5.

1. DO's and Hecke Operators

In this section we collect together basic facts about ψDO's and Hecke operators.

Γ will always denote a discrete subgroup of $SL_2(\mathbb{R})$ so that $\Gamma\backslash\mathfrak{h}$ has finite area ("cofinite subgroup"). Its commensurator $\mathrm{Comm}(\Gamma)=\{g\in SL_2(\mathbb{R})\colon [\Gamma\colon \Gamma'(g)]<\infty$ and $[g^{-1}\Gamma g\colon \Gamma'(g)]<\infty\}$, where $\Gamma'(g)=g^{-1}\Gamma g\cap\Gamma$. If $g\in\mathrm{Comm}(\Gamma)$ and $g\notin\Gamma$, one has a Hecke Operator T_g.

Definition 1.1. Let $\Gamma=\bigcup_{j=1}^{d}\Gamma'(g)\gamma_j$ (disjoint). Then $T_g f(x)=\sum_{j=1}^{d} f(g\gamma_j x)$.

We will need the following information about the T_g.

Proposition 1.1. (i) $\Gamma=\bigcup_{j=1}^{d}\Gamma'(g)\gamma_j$ *iff* $\Gamma g\Gamma=\bigcup_{j=1}^{d}\Gamma g\gamma_j$,

(ii) $T_g^*=T_{g^{-1}}$.

(iii) *There exists* $\alpha_1,\ldots,\alpha_d$ *so that* $\Gamma g\Gamma=\bigcup_{j=1}^{d}\Gamma\alpha_j=\bigcup_{j=1}^{d}\alpha_j\Gamma$ *(all unions are always disjoint).*

Proof. (i) See [H, Proposition 2.7] or [S].

(ii) See [H, Proposition 2.16] or [S].

(iii) We will need to use the proof of this, so we recall it here (cf. [S Lemma 3.5]). First, let $\Gamma\alpha\Gamma=\bigcup_{j=1}^{d}\Gamma\alpha_j=\bigcup_{j=1}^{e}\beta_j\Gamma$ be disjoint coset decompositions. Then $d=e$ ([S, pp. 53–54]). So $\alpha_j\in\Gamma\beta_k\Gamma$ for any k. Pair up the $\{\alpha_j\}$ and $\{\beta_j\}$ in some fixed way, and reindex so α_j is paired to β_j. Then $\alpha_j=\delta_j\beta_j\varepsilon_j$ where $\delta_j,\varepsilon_j\in\Gamma$. Let $\tilde\alpha_j=\delta_j^{-1}\alpha_j$. Then $\bigcup_{j=1}^{d}\Gamma\tilde\alpha_j=\bigcup_{j=1}^{d}\tilde\alpha_j\Gamma$. ∎

We will henceforth assume coset representatives $\{\alpha_j\}$ to be two-sided as in (iii). Then $T_g f(x)=\sum_{j=1}^{d} f(\alpha_j x)$ and $T_g^* f(x)=\sum_{j=1}^{d} f(\alpha_j^{-1}x)$.

Let us also use the notation $g\cdot\psi$ for the left translate of a function ψ: $g\cdot\psi(x)=\psi(gx)$. Thus we will write $T_g=\sum_{j=1}^{d}\alpha_j$, e.g.

Next we recall definition and notation for ψDO's on $\Gamma\backslash\mathfrak{h}$. For $z\in\mathfrak{h}$, $b\in\mathbb{R}=bd\,\mathfrak{h}$, let $\xi(z,b)$ be the horocycle through z and b, and let $\langle z,b\rangle$ be the (least) distance of $\xi(z,b)$ to i. Then $e_{\lambda,b}(z)=e^{(i\lambda+1)\langle z,b\rangle}$ are a complete family of eigenfunctions for Δ on $L^2(D)$. Given $a\in C^\infty(\mathfrak{h}\times B\times\mathbb{R}^+)$ we define $\mathrm{Op}(a)$ by the equation: $e^{-(i\lambda+1)\langle z,b\rangle}\,\mathrm{Op}(a)\,e^{(i\lambda+1)\langle z,b\rangle}=a(z,b,\lambda)$. $\mathrm{Op}(a)$ is a ψDO of order m if $a\sim\sum_{j=m}^{-\infty}a_j(z,b,\lambda)$ with $a_j(z,b,\lambda)\sim\lambda^j a_j(z,b,1)$ as $\lambda\to\infty$ and with $\sim$ in the sense of the symbol norms (cf. [Tr, Z1]). $Op(a)$ defines a ψDO on $\Gamma\backslash\mathfrak{h}$ iff $\gamma\,\mathrm{Op}(a)=\mathrm{Op}(a)\,\gamma$ for all $\gamma\in\Gamma$ iff $a(\gamma z,\gamma b,\lambda)=a(z,b,\lambda)$. One has an "explicit" formula for $\mathrm{Op}(a)$ on eigenfunctions φ_k of Δ on $\Gamma\backslash\mathfrak{h}$. Indeed, according to Helgason's eigenfunction representation theorem:

$$\varphi_k(z)=\int_B e^{(i\lambda_k+1)\langle z,b\rangle}\,dT_k(b)$$

if φ_k is an eigenfunction. Consequently

$$\mathrm{Op}(a)\,\varphi_k(z) = \int_B a(z, b, \lambda_k)\, e^{(i\lambda_k+1)\langle z, b\rangle}\, dT_k(b).$$

More generally, we will be considering certain special kinds of Fourier Integral Operators (FIO's). These FIO's are again characterized by their effect on the $\{e^{(i\lambda+1)\langle z,b\rangle}\}$: $Ue^{(i\lambda+1)\langle z,b\rangle}$ will be a finite linear combination of terms $a(z, b, \lambda)\, e^{i\varphi(z, b, \lambda)}$, where a is a symbol of order 0 and φ is homogeneous of degree 1 in λ. (In fact, all our FIO's are just sums of compositions of left translations and ψDO's.) We will refer to $e^{-(i\lambda+1)\langle z,b\rangle}\, Ue^{(i\lambda+1)\langle z,b\rangle}$ as the *amplitude of U*. Again, it is Γ-bi-invariant if and only if $U\gamma = \gamma U$ for all $\gamma \in \Gamma$.

Finally $(\Gamma\backslash\mathfrak{h} \times B) \times \mathbb{R}^+$ is identified with $T(\Gamma\backslash\mathfrak{h} \times B)$ by the map taking (z, b, λ) to the vector of length λ tangent to the geodesic through (z, b) (cf. [Z1]).

2. Computation of Operator Amplitudes

In this section we compute the amplitude of the operator $(T_g\,\mathrm{Op}(a)\,T_g^*)^m\,\mathrm{Op}(\sigma)$. Let us denote by g the operator of left translation by g: $g\psi(z) = \psi(gz)$.

Proposition 2.1. $g_1\,\mathrm{Op}(a)\,g_2 e^{(i\lambda+1)\langle z,b\rangle} = a(g_1 z, g_2^{-1} b, \lambda)\, e^{(i\lambda+1)\langle g_2 g_1 z, b\rangle}$.

Proof. From the identity $\langle gz, gb\rangle = \langle z, b\rangle + \langle g\cdot 0, g\cdot b\rangle$ (cf. [He]), we have

$$\begin{aligned}\mathrm{Op}(a)\, g_2\, e^{(i\lambda+1)\langle z,b\rangle} &= \mathrm{Op}(a)(e^{(i\lambda+1)\langle z, g_2^{-1} b\rangle}\, e^{(i\lambda+1)\langle g_2\cdot O, g_2\cdot b\rangle}) \\ &= a(z, g_2^{-1} b, \lambda)\, e^{(i\lambda+1)\langle g_2 z, b\rangle}.\end{aligned}$$

Left translating by g_1 gives $a(g_1 z, g_2^{-1} b)\, e^{(i\lambda+1)\langle g_2 g_1 z, b\rangle}$. ∎

Corollary 2.1. *$g\,\mathrm{Op}(a)\,g^{-1}$ is a ψDO with complete symbol $a(gz, gb, \lambda)$.*

Now $T_g\,\mathrm{Op}(a)\,T_g^* = \sum_{i,j} \alpha_j\,\mathrm{Op}(a)\,\alpha_i^{-1} = \sum_i \alpha_i\,\mathrm{Op}(a)\,\alpha_i^{-1} + \sum_{i\neq j} \alpha_j\,\mathrm{Op}(a)\,\alpha_i^{-1}$. By Corollary 2.1, the first term is $\mathrm{Op}(T_g a)$. Thus,

Corollary 1.3. *$T_g\,\mathrm{Op}(a)\,T_g^* = \mathrm{Op}(T_g a) + R_g(a)$, where $R_g(a)$ is an* FIO *with amplitude $\sum_{i\neq j} a(\alpha_j z, \alpha_i b)\, e^{-(i\lambda+1)\langle z,b\rangle}\, e^{(i\lambda+1)\langle \alpha_i^{-1}\alpha_j z, b\rangle}$.*

To get the compositions $(T_g\,\mathrm{Op}(a)\,T_g^*)^m\,\mathrm{Op}(\sigma)$, we use the standard:

PROPOSITION 2.2. *Let a, σ be symbols of order* 0, *and* $g \in SL_2(\mathbb{R})$. *Then* $\mathrm{Op}(a)(\sigma e^{(i\lambda+1)\langle z, gb\rangle}) = a(z, gb)\,\sigma(z, b)\, e^{(i\lambda+1)\langle z, gb\rangle} + O(\lambda^{-1})$.

Proof. This just requires translating into our notation the 0th order part of the "fundamental expansion lemma of FIO theory" [T]. Let (x, ξ) denote the symplectic coordinates in T^*X. Then if $\lambda(x, \xi)$ and $\sigma_1(x, \xi)$ are 0th order symbols, and $\varphi(x, \xi)$ is a first order symbol, $\lambda(x, D_x)(\sigma_1 e^{i\varphi}) = \sigma_2 e^{i\varphi}$, where $\sigma_2(x, \xi) = \sigma_1(x, \xi)\,\lambda(x, \nabla\varphi) + O(|\xi|^{-1})$. In our $\mathfrak{h} \times B \times \mathbb{R}^+$ coordinates (z, b, λ), φ is $\lambda\langle z, gb\rangle$ while σ_1 is $\sigma(z, b)\, e^{\langle z, gb\rangle}$. Moreover, $\nabla\langle z, gb\rangle$ is the unit vector along the geodesic from z to gb, which in our coordinates is (z, gb). Thus $\mathrm{Op}(a)(\sigma e^{\langle z, gb\rangle}\, e^{i\lambda\langle z, gb\rangle})$ is $a(z, gb)\,\sigma(z, b)\, e^{(i\lambda+1)\langle z, gb\rangle}$. ∎

Repeated use of Corollary 1.3 and Proposition 2.2 gives

PROPOSITION 2.3. $(T_g\,\mathrm{Op}(a)\,T_g^*)^m\,\mathrm{Op}(\sigma)\, e^{(i\lambda+1)\langle z, b\rangle} = \sum_{i_1, \ldots, i_m, j_1, \ldots, j_m} A(\sigma, \mathbf{i}, \mathbf{j}, z, b)\; e^{(i\lambda+1)\langle \alpha_{i_1}^{-1}\alpha_{j_1}\cdots\alpha_{i_m}^{-1}\alpha_{j_m} z, b\rangle}$, where

$$\begin{aligned} &A(\sigma, \mathbf{i}, \mathbf{j}, z, b)\\ &\quad = a(\alpha_{j_m} z, (\alpha_{i_1}^{-1}\alpha_{j_1}\cdots\alpha_{i_m}^{-1})^{-1} b)\, a(\alpha_{j_{m-1}}\alpha_{i_m}^{-1}\alpha_{j_m} z, (\alpha_{i_1}^{-1}\cdots\alpha_{m-1}^{-1})^{-1} b)\cdots\\ &\quad\quad \cdots \times a(\alpha_{j_1}\alpha_{i_2}^{-1}\alpha_{j_2}\cdots\alpha_{i_m}^{-1}\alpha_{j_m} z, \alpha_{i_1} b)\,\sigma(\alpha_{i_1}^{-1}\alpha_{j_1}\cdots\alpha_{i_m}^{-1}\alpha_{j_m} z, b). \end{aligned}$$

3. TRACE FORMULA AND MOMENTS

In this section, we give a preliminary formula for the moments $\int x^m\, d\mu_{g,a}(x) = \lim_{\lambda\to\infty} (1/N(\lambda))\,\mathrm{tr}(\pi_\lambda T_g\,\mathrm{Op}(a)\,T_g^*\,\pi_\lambda)^m$ of the Szegö limit distribution $\mu_{g,a}$. We first follow Widom's argument in [Wi] to equate this to $\mathrm{tr}_\Delta(T_g\,\mathrm{Op}(a)\,T_g^*)^m$. We then briefly review Sarnak's argument in [Sa] to evaluate the latter.

PROPOSITION 3.1. *For any arithmetic* Γ,

$$\lim_{\lambda\to\infty}\frac{1}{N(\lambda)}\,\mathrm{tr}(\pi_\lambda T_g\,\mathrm{Op}(a)\,T_g^*\,\pi_\lambda)^m = \lim_{\lambda\to\infty}\frac{1}{N(\lambda)}\,\mathrm{tr}\,\pi_\lambda(T_g\,\mathrm{Op}(a)\,T_g^*)^m\,\pi_\lambda.$$

Proof. We must show $\mathrm{tr}[\pi_\lambda(T_g\mathrm{Op}(a)\,T_g^*)^m\pi_\lambda - (\pi_\lambda T_g\mathrm{Op}(a)\,T_g^*\,\pi_\lambda)^m] = o(N(\lambda))$. As pointed out in [Wi], this follows if and only if $\|\pi_\lambda(T_g\,\mathrm{Op}(a)\,T_g^*)(I-\pi_\lambda)\|_2^2 = o(N(\lambda))$ where $\|\cdot\|_2$ is the Hilbert–Schmidt norm. Only one novelty arises in applying Widom's estimate of this norm. Widom writes $\|\pi_\lambda(T_g(\mathrm{Op}(a)\,T_g)(I-\pi_\lambda)\|_2^2 = \|\pi_\lambda(T_g\mathrm{Op}(a)\,T_g^*)(\pi_{\lambda+\delta}-\pi_\lambda)\|_2^2 + \|\pi_\lambda(T_g\,\mathrm{Op}(a)\,T_g^*)(I-\pi_{\lambda+\delta})\|_2^2$ for arbitrary δ. For any arithmetic Γ, the first term is $O(N(\lambda+\delta)-N(\lambda)) = o(N(\lambda))$. For the second, one uses the identity $\pi\{\lambda_i\}\,T_g\,\mathrm{Op}(a)\,T_g^*\,\pi\{\lambda_j\} = (1/(\lambda_i-\lambda_j))\;\pi\{\lambda_i\}$

$[T_g \operatorname{Op}(a) T_g^*, \sqrt{\Delta}]\, \pi\{\lambda_j\}$, where $\pi\{\lambda_j\}$ is the spectral projection for the eigenvalue λ_j. The point is that $[T_g \operatorname{Op}(a) T_g^*, \sqrt{\Delta}] = T_g[\operatorname{Op}(a), \sqrt{\Delta}]\, T_g^*$ is a bounded FIO. So we can follow Widom in estimating the second term as $O(\delta^{-2} N(\lambda))$. Since δ is arbitrary, $\|\pi_\lambda(T_g(\operatorname{Op}(a) T_g^*)(I-\pi_\lambda)\|_2^2 = o(N(\lambda))$. ∎

The above proposition reduces the calculation of the moments to a trace, and naturally one uses the trace formula. We will refer to [L–P, C–S, Sa] for background in the trace formulae. The main assumption we make is again that Γ is arithmetic. It follows that the continuous spectrum has zero density in the spectrum, and that $\operatorname{tr}_\Delta A$ is the coefficient of the big singularity at $t=0$ of $\operatorname{tr}\, \pi_d \cos t\sqrt{-(\Delta+1/4)}\, A$ where π_d is projection onto the discrete subspace.

THEOREM 3.3. *Suppose Γ arithmetic. Then*

$$\operatorname{tr}_\Delta(T_g \operatorname{Op}(a) T_g^*)^m \operatorname{Op}(\sigma)$$

$$= \sum_{\substack{(i_1,\dots,i_m)\\(j_1,\dots,j_m)\\ \alpha_{i_1}^{-1}\alpha_{j_1}\cdots\alpha_{i_m}^{-1}\alpha_{j_m}\in\Gamma}} \iint_{F\times B} A(\sigma, \mathbf{i}, \mathbf{j}, z, b)\, e^{2\langle \alpha_{i_1}^{-1}\alpha_{j_1}\cdots\alpha_{i_m}^{-1}\alpha_{j_m} z, b\rangle}\, d\operatorname{vol}(z)\, db,$$

where F is a fundamental domain for Γ.

Remark. $e^{2\langle z,b\rangle}\, d\operatorname{vol}(z)\, db$ is Haar measure (or Liouville measure).

Proof. We follow Sarnak's argument, and refer to [Sa] for the full details.

It is sufficient to compute the coefficient of the main singularity at $t=0$ of $\operatorname{tr}\, \pi_d \cos t\sqrt{L}[T_g \operatorname{Op}(a) T_g^*)^m \operatorname{Op}(\sigma)$, where $L = -(\Delta+\frac{1}{4})$. To find this singularity we need the kernel of $(T_g \operatorname{Op}(a) T_g^*)^m \operatorname{Op}(\sigma) \cos t\sqrt{L}$ on $\mathfrak{h}$. This kernel, $K(g, a, \sigma; t, z, w)$ is gotten by applying $(T_g \operatorname{Op}(a) T_g^*)^m \operatorname{Op}(\sigma)$ to $\cos t\sqrt{L}(z,w) = \int_0^\infty \int_B e^{(i\mu+1)\langle z,b\rangle}\, e^{(-i\mu+1)\langle w,b\rangle} \cos t\mu \,d\!\!\!/\mu\, db$, where $d\!\!\!/\mu = \mu\tanh(\pi\mu/2)$ and db is Lebesgue measure. We have computed $(T_g \operatorname{Op}(a) T_g^*)^m \operatorname{Op}(\sigma)\, e^{(i\mu+1)\langle z,b\rangle}$ up to terms of order (μ^{-1}), which cause a lower order singularity. We also drop the lower order terms of a and σ, and keep only the homogeneous parts of order 0, a_0, and σ_0. Writing the resulting amplitude as A_0, we find that it's enough to consider the kernel $K_0(g, a, \sigma; t, z, w) = \sum_{(i_1,\dots,i_m),(j_1,\dots,j_m)} \int_B \int_0^\infty A_0(\sigma, \mathbf{i}, \mathbf{j}, z, b)$ $e^{(i\mu+1)\langle \alpha_{i_1}^{-1}\alpha_{j_1}\cdots\alpha_{i_m}^{-1}\alpha_{j_m} z, b\rangle}\, e^{(-i\mu+1)\langle w,b\rangle} \cos t\mu \,d\!\!\!/\mu\, db$. The kernel on $\Gamma\backslash\mathfrak{h}$ is then $k_0(g, a, \sigma; t, z, w) = \sum_{\gamma\in\Gamma} K_0(g, a, \sigma, t, z, \gamma w)$. As the continuous spectral part is of lower order, we have that the main singularity of $\operatorname{tr}\, \pi_d \cos t\sqrt{L}(T_g \operatorname{Op}(a) T_g^*)^m \operatorname{Op}(\sigma)$ is identical to that of $\int_F k_0(g, a, \sigma, t, z, z)\, dvol(z)$, where F is a fundamental domain.

Now

$$\int_F k_0(g, a, \sigma; t, z, z)\, dvol(z)$$

$$= \sum_{\substack{(i_1, \ldots, i_m)\\(j_1, \ldots, j_m)}} \sum_{\gamma \in \Gamma} \int_F \int_B \int_O^\infty A_0(\sigma, \mathbf{i}, \mathbf{j}, z, b)\, e^{(i\mu+1)\langle \alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m} z, b\rangle}$$

$$\times e^{(-i\mu+1)\langle \gamma z, b\rangle} \cos t\mu\, d\mu db dvol(z)$$

The singularities of the trace occur when the phases $\mu(\langle \alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m} z, b\rangle - \langle \gamma z, b\rangle \pm t)$ are stationary. As discussed in [Sa], the main singularity occurs when $t = 0$ and $\alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m} \in \Gamma$. This gives $\int_0^\infty \cos t\mu d\mu$ times the coefficient $\sum_{(i_1, \ldots, i_m), (j_1, \ldots, j_m), \alpha_{i_1}^{-1}\alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m} \in \Gamma}$ $\int_F \int_B A_0(\sigma, \mathbf{i}, \mathbf{j}, z, b)\ e^{2\langle \alpha_{i_1}^{-1}\alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m} z, b\rangle}\, db dvol(z)$. This coefficient is then $\operatorname{tr}_\Delta(T_g \operatorname{Op}(a)\, T_g^*)^m \operatorname{Op}(\sigma)$. ∎

COROLLARY 3.4. $\operatorname{tr}_\Delta(T_g \operatorname{Op}(a)\, T_g^*) \operatorname{Op}(\sigma) = \int_F \int_B (T_g a_0)(z, b)\, \sigma_0(z, b)\, d\omega$.

Proof. For $m = 1$ it is clear that $\alpha_{i_1}^{-1} \alpha_{j_1} \in \Gamma$ iff $\alpha_{i_1}^{-1} \alpha_{j_1} = e$ since cosets were disjoint. We thus get $\sum_i \int_F \int_B a(\alpha_i z, \alpha_i b)\, \sigma(z, b)\, d\omega$. ∎

We now need to analyze the coefficient for $m \geqslant 2$. This requires analyzing the condition $\alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m} \in \Gamma$, which we take up in the next section.

4. SPELLING ELEMENTS OF $SL_2(\mathbb{Z})$

In this section we determine exactly when the condition $\alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m} \in SL_2(\mathbb{Z})$ is satisfied by a specific set of two-sided coset representatives for the double cosets $\Gamma\left(\begin{smallmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2} \end{smallmatrix}\right) \Gamma$, p a prime.

To begin with we need some basic facts about $\Gamma\left(\begin{smallmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2} \end{smallmatrix}\right) \Gamma$. See,e.g., [S] for elementary background.

PROPOSITION 4.1. *The following elements* $\{\alpha_j\}$ *of* $SL_2(\mathbb{R}) = \Gamma_1$ provide two sided coset representatives for $\Gamma_1\left(\begin{smallmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2} \end{smallmatrix}\right) \Gamma_1$,

$$\alpha_j = \begin{pmatrix} -jp^{-1/2} & -p^{1/2} - j^2 p^{-1/2} \\ p^{-1/2} & jp^{-1/2} \end{pmatrix}, \qquad 0 \leqslant j \leqslant p-1, \alpha_p = \begin{pmatrix} 0 & p^{-1/2} \\ -p^{1/2} & 0 \end{pmatrix}.$$

Proof. (i) $\Gamma_1\left(\begin{smallmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2} \end{smallmatrix}\right) \Gamma_1 = p^{-1/2} \Gamma_1 \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right) \Gamma_1 = p^{-1/2}\left(\bigcup_{j=0}^{p-1} \Gamma_1 \left(\begin{smallmatrix} 0 & p \\ -1 & -j \end{smallmatrix}\right) \cup \Gamma_1 \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)\right)$.

(ii) $\Gamma_1\left(\begin{smallmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2}\end{smallmatrix}\right)\Gamma_1 = \left(\Gamma_1\left(\begin{smallmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2}\end{smallmatrix}\right)\Gamma_1\right)^{-1} \;=\; p^{-1/2}\left(\bigcup_{j=0}^{p-1}\left(\begin{smallmatrix} -j & -p \\ 1 & 0\end{smallmatrix}\right)\Gamma_1 \cup \left(\begin{smallmatrix} 1 & 0 \\ 0 & p\end{smallmatrix}\right)\Gamma_1\right)$.

(iii) $p^{-1/2}\left(\begin{smallmatrix} 0 & p \\ -1 & -j\end{smallmatrix}\right) = p^{-1/2}\left(\begin{smallmatrix} -1 & -j \\ 0 & -1\end{smallmatrix}\right)\left(\begin{smallmatrix} -j & -p \\ 1 & 0\end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & j \\ 0 & 1\end{smallmatrix}\right)$.

$$p^{-1/2} w \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} w^{-1} = p^{-1/2}\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}, \qquad \text{where} \quad w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

(iv) Following Proposition 1.1(iii), we get

$$\alpha_j = p^{-1/2}\begin{pmatrix} -1 & -j \\ 0 & -1 \end{pmatrix}^{-1}\begin{pmatrix} 0 & p \\ -1 & -j \end{pmatrix} = p^{-1/2}\begin{pmatrix} -j & -p-j^2 \\ 1 & j \end{pmatrix}, \qquad 0 \leqslant j \leqslant p-1$$

$$\alpha_p = p^{-1/2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} = p^{-1/2}\begin{pmatrix} 0 & 1 \\ -p & 0 \end{pmatrix}.$$

Then the α_j are two sided double coset representatives. ∎

Remark that $\{\alpha_j, \alpha_p\}$ are all reflections in $PSL_2(\mathbb{R})$. Let $\langle \alpha_0, \alpha_1, \ldots, \alpha_p\rangle$ be the group they generate.

PROPOSITION 4.2. (i) *$\langle \alpha_0, \alpha_1, \ldots, \alpha_p\rangle$ is a free product of reflections $\langle \alpha_i \rangle$.*

(ii) *$\alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_m} \in PSL_2(\mathbb{Z})$ iff it spells the identity.*

Proof. Sarnak has shown that no word of odd length in the α_i's can be in $PSL_2(\mathbb{Z})$. So we need only show that no even product is in $PSL_2(\mathbb{Z})$ unless it spells the identity e, and then iff it reduces to e by cancellation. This we do as follows.

It is clear that $\alpha_{i_1}\cdots\alpha_{i_{2m}} \in PSL_2(\mathbb{Z})$ only if its term of order p^{-m} is $\equiv O(p)$. Now $\alpha_j = p^{-1/2}\left(\begin{smallmatrix} -j & -j^2 \\ 1 & j\end{smallmatrix}\right) + O(p^{1/2})$, and $\alpha_p = p^{-1/2}\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0\end{smallmatrix}\right) + O(p^{1/2})$. The top term arises if we drop terms of order $O(p^{1/2})$. So, letting $\hat{\alpha}_j = \left(\begin{smallmatrix} -j & -j^2 \\ 1 & j\end{smallmatrix}\right)$ and $\hat{\alpha}_p = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0\end{smallmatrix}\right)$, we will show that $\hat{\alpha}_{i_1}\cdots\hat{\alpha}_{i_{2m}} \equiv O(p)$ iff $\alpha_{i_1}\cdots\alpha_{i_{2m}}$ reduces to e.

It is convenient to work with products of two. These are:

(a) $\hat{\alpha}_{jk} = \hat{\alpha}_j\hat{\alpha}_k = (j-k)\left(\begin{smallmatrix} -j & -jk \\ 1 & k\end{smallmatrix}\right)$ (type I: $0 \leqslant j, k \leqslant p-1$),

(b) $\hat{\alpha}_{jp} = \hat{\alpha}_j\hat{\alpha}_p = \left(\begin{smallmatrix} 0 & -j \\ 0 & 1\end{smallmatrix}\right)$ (type II: $0 \leqslant j \leqslant p-1$),

(c) $\hat{\alpha}_{pj} = \hat{\alpha}_p\hat{\alpha}_j = \left(\begin{smallmatrix} 1 & j \\ 0 & 0\end{smallmatrix}\right)$ (type III: $0 \leqslant j \leqslant p-1$).

Thus we must consider m-fold product of the $\hat{\alpha}$'s. Successive cancellation occurs when two adjacent indices are equal. Thus we henceforth assume: no adjacent indices are equal.

The key to computing large products is that products of two of the above $\hat{\alpha}$'s is just another $\hat{\alpha}$, with one exception. We will write the product in terms of types. First, type I $\cdot$ I = I: $\hat{\alpha}_{jk}\,\hat{\alpha}_{ml} = (m-l)(j-k)(k-m)\left(\begin{smallmatrix} -j & -jl \\ 1 & l \end{smallmatrix}\right)$. We abbreviate as $\hat{\alpha}_{jk}\,\hat{\alpha}_{ml} \propto \hat{\alpha}_{jl}$, where the proportionality constant is never divisible by p.

It is clear that no nontrivial product of type I $\hat{\alpha}$'s can be $\equiv O(p)$. So we can conclude that $\langle \alpha_0, \ldots, \alpha_{p-1} \rangle$ is a free product of reflections. We now turn to the cases where α_p occurs:

$$\text{Type I} \cdot \text{II} = \text{II}: \hat{\alpha}_{jk}\hat{\alpha}_{mp} = (j-k)(k-m)\begin{pmatrix} 0 & -j \\ 0 & 1 \end{pmatrix} \propto \hat{\alpha}_{jp}(p \nmid (j-k)(k-m))$$

$$\text{Type II} \cdot \text{I} = \text{I}: \hat{\alpha}_{mp}\hat{\alpha}_{jk} = (j-k)\begin{pmatrix} -m & -mk \\ 1 & k \end{pmatrix} \propto \hat{\alpha}_{mk}(p \nmid (j-k))$$

$$\text{Type II} \cdot \text{II} = \text{II}: \alpha_{jp}\alpha_{mp} = \alpha_{jp}$$

$$\text{Type I} \cdot \text{III} = \text{I}: \hat{\alpha}_{jk}\hat{\alpha}_{pm} = (j-k)\begin{pmatrix} -j & -jm \\ 1 & m \end{pmatrix} \propto \hat{\alpha}_{jm}\,(p \nmid (j-k))$$

$$\text{Type III} \cdot \text{I} = \text{III}: \hat{\alpha}_{pm}\hat{\alpha}_{jk} = (j-k)(m-j)\begin{pmatrix} 1 & k \\ 0 & 0 \end{pmatrix} \propto \hat{\alpha}_{pk}\,(p \nmid (j-k)(m-j))$$

$$\text{Type III} \cdot \text{II}: \hat{\alpha}_{pj}\hat{\alpha}_{mp} = (j-m)\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \text{the exception}$$

$$\text{Type III} \cdot \text{III} = \text{III}: \hat{\alpha}_{pj}\hat{\alpha}_{pm} = \hat{\alpha}_{pm}.$$

An easy conclusion: no unreduced word can spell element of $SL_2(\mathbb{Z})$ unless a factor of type III $\cdot$ II appears in it. More generally, consider a product $\hat{\alpha}_{i_1,j_1} \cdots \hat{\alpha}_{i_m,j_m}$. Two products of type III $\cdot$ II cannot occur adjacently (else the word is reducible). So products of type III $\cdot$ II must alternate with factors of type I, II, or III, i.e., the word is $*\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)*\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)* \cdots \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)*$ (the first or last $*$'s might not occur). We argue inductively that no such product is $\equiv O(p)$ unless it is reducible. The cases $m = 1, 2$ are clear by the computation above. So let $m \geqslant 3$. We observe that if a $*$ occurs to the left of a $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, it cannot be of type II. Indeed, an $\hat{\alpha}_{jp}$ of type II cannot occur adjacent to $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ in the original word (else it's reducible), and $*$ can be of type II only if the right-most factor of it in the original is of type II. Similarly a $*$ occuring to the right of a $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ cannot be of type III (the left-most factor must have been type III). Thus an irreducible product must have the form $\left(\begin{smallmatrix} \text{I} \\ \text{III} \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)(\text{I})\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \cdots \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} \text{I} \\ \text{II} \end{smallmatrix}\right)$, where the first or last factor may be absent. Actually III $\cdot \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)(\text{II}) = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, so these cases can be omitted, too. Finally, $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} -j & -jk \\ 1 & k \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & k \\ 0 & 0 \end{smallmatrix}\right) = (\text{III})$, and $\left(\begin{smallmatrix} -j & -jk \\ 1 & k \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & -j \\ 0 & 1 \end{smallmatrix}\right) = (\text{II})$. We are reduced to a smaller product $*\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)*\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)* \cdots$, and the argument is complete by induction. ∎

5. Szegö Limit Theorems

We now have all the elements required to prove the main theorem of this paper. Recall that a self-adjoint operator A is said to have its eigenvalues distributed according to the Szegö limit measure μ_A on $\mathbb{R}$ if $\int_{-\infty}^{\infty} x^m \, d\mu_A(x) = \lim_{\lambda \to \infty} (1/N(\lambda)) \operatorname{tr}(\pi_\lambda A \pi_\lambda)^m$ (π_λ are orthoprojections onto the span of eigenfunctions of Δ of eigenvalue $\leqslant \lambda$).

THEOREM 5.1. *Let* $\mathrm{Op}(a)$ *be a* 0-th *order* ψDO *on* $L^2(\Gamma_1 \backslash \mathfrak{h})$, $\Gamma_1 = SL_2(\mathbb{Z})$, *and let* T_p *be the Hecke Operator corresponding to the double coset* $\Gamma_1 \left(\begin{smallmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2} \end{smallmatrix}\right) \Gamma_1$. *Then the Szegö limit measure* $\mu_{a,p}$ *of* $T_p \, \mathrm{Op}(a) \, T_p^*$ *has the moments*

$$\int x^m \, d\mu_{a,p}(x)$$

$$= \int_F \sum_{\substack{(i_1, \ldots, i_m) \\ (j_1, \ldots, j_m) \\ \alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m}}} a_0(\alpha_{j_m} \xi) \, a_0(\alpha_{j_{m-1}} \alpha_{i_1}^{-1} \alpha_{j_m} \xi) \cdots a_0(\alpha_{j_1} \alpha_{i_1}^{-1} \cdots \alpha_{j_m} \xi) \, d\omega,$$

where F *is a fundamental domain for* Γ_1 *in* $SL_2(\mathbb{R})$, $d\omega$ *is Haar measure, and the* $\alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m}$ *fully reduces to* e.

Proof. Combine Proposition 3.1, Theorem 3.3, and Proposition 4.2. ∎

Recall that $T_p \, \mathrm{Op}(a) \, T_p^* = \mathrm{Op}(T_p a) + R_p(a)$. The same proof allows us to determine the Szegö limit distribution for $R_p(a)$.

THEOREM 5.2. *The Szegö limit measure* $d\mu_{R_p(a)}$ *for* $R_p(a)$ *has the moments*

$$\int x^m \, d\mu_{R_p(a)} = \int_F \sum_{\substack{i_1 \neq j_1 \\ i_m \neq j_m \\ \alpha_{i_1}^{-1} \alpha_{j_1} \cdots \alpha_{i_m}^{-1} \alpha_{j_m} = e}} a(\alpha_{j_m} \xi) \cdots a(\alpha_{j_1} \alpha_{i_1}^{-1} \cdots \alpha_{j_m} \xi) \, d\omega(\xi)$$

(*same notation as before*).

Proof. Since $[R_p(a), \sqrt{\Delta}]$ is bounded, we have by Proposition 3.1 that $\int x^m \, d\mu_{R_p(a)} = \operatorname{tr}_\Delta R_p(a)^m$. But $\operatorname{tr}_\Delta R_p(a)^m = \operatorname{tr}_\Delta \sum_{i_1 \neq j_1, i_m \neq j_m} \alpha_{j_m} \mathrm{Op}(a) \, \alpha_{i_m}^{-1} \cdots \alpha_{j_1} \mathrm{Op}(a) \, \alpha_{i_1}^{-1}$, and this can be computed exactly as in Theorem 3.3. ∎

Of course one has $\operatorname{tr}_\Delta R_p(a)^m \, \mathrm{Op}(\sigma) = \int_F (\sum \ldots) \, \sigma(\xi) \, d\omega$, where $(\sum \ldots)$ is the same as in Theorem 5.2. Using these formulae, it is straightforward to give a closed formula for a given moment of $d\mu_{a,p}$ or $d\mu_{R_p(a)}$ in terms of T_p and a. For example:

COROLLARY 5.3.

$$(m=2) \qquad \int x^2\, d\mu_{a,p} = \int_F ((T_p a)^2 + pT_p(a^2))\, d\omega,$$

$$(m=3) \qquad \int x^3\, d\mu_{a,p} = \int_F ((T_p a)^3 + 3pT_p(a^2)\, T_p(a) + p(p-1)\, T_p(a^3))\, d\omega.$$

Proof.

$$(m=2) \qquad \int x^2 d\mu_{a,p} = \operatorname{tr}_\Delta(\operatorname{Op}(T_p a) + R_p(a))^2; \operatorname{tr}_\Delta \operatorname{Op}(T_p\, a)^2 = \int_F (T_p\, a)^2\, d\omega; \operatorname{tr}_\Delta R_p(a) \operatorname{Op}(T_p\, a) = 0;$$

$$\operatorname{tr}_\Delta R_p(a)^2 = \int_F pT_p(a^2)\, d\omega.$$

$$(m=3) \qquad \int x^3\, d\mu_{a,p} = \operatorname{tr}_\Delta(\operatorname{Op}(T_p a)^3 + 3\operatorname{Op}(T_p a)^2 R_p(a) + 3R_p(a)^2 \operatorname{Op}(T_p a) + R_p(a)^3)$$
$$= \int_F ((T_p\, a)^3 + 3pT_p(a)\, T_p(a^2) + p(p-1)\, T_p(a^3))\, d\omega.$$

OPEN PROBLEMS AND DISCUSSIONS

(i) The higher moments can doubtless also be expressed in terms of T_p and a. However, the combinatorics involved appear quite complicated, and we do not know a closed formula for the mth moment; lakcing that, we have no proof that all moments can be expressed in this way.

(ii) Is the result in Section 4, that the $\{\alpha_j\}$ generate a coxeter group, true for Quaternion groups?

(iii) Following [Z2], let $\{d\mathscr{U}_k\} \subset \mathscr{D}'(\Gamma\backslash PSL_2(\mathbb{R}))$ be defined by: $(\operatorname{Op}(a)\, u_k, u_k) = \int a\, d\mathscr{U}_k$, where the $\{u_k\}$ are, as above, the simultaneous eigenfunctions of Δ and of the T_p. The formulae

$$(\operatorname{Op}(a)\, u_k, u_k) = (e^{-it\sqrt{\Delta}} \circ \operatorname{Op}(a)\, e^{it\sqrt{\Delta}}\, u_k, u_k) = (\operatorname{Op}(a \circ G^t)\, u_k, u_k) + R_t(a)$$

(with G^t = geodesic flow, R_t compact) imply that the weak* limit measures of the $\{dU_k\}$ are G^t-invariant. What invariance properties of these limit measures follow from the formula:

$$(T_p \operatorname{Op}(a)\, T_p^* u_k, u_k) = |\rho_k(p)|^2 (\operatorname{Op}(a)\, u_k, u_k)?$$

REFERENCES

[C–S] P. COHEN AND P. SARNAK, Discrete groups and harmonic analysis, preprint, 1986.

[H] D. HEJHAL, "The Selberg Trace Formula for $PSL_2(\mathbb{R})$," Vols. 1 and 2, Lecture Notes, Vols. 548, 1001, Springer, Berlin, 1976, 1983.

[He] S. HELGASON, "Topics in Harmonic Analysis on Homogeneous Spaces," Progress in Math., Vol. 13, Birkhauser, Basel, 1981.

[L–P] P. LAX AND R. PHILLIPS, "Scattering Theory for Automorphic Functions," Annals of Math. Studies, No. 87, Princeton Univ. Press, Princeton, NJ, 1976.

[Sa] P. Sarnak, The Sato–Tate conjecture, preprint.

[S] G. SHIMURA, "Introduction to the Arithmetic Theory of Automorphic Functions," Princeton Univ. Press, Princeton, NJ, 1971.

[T] M. TAYLOR, Pseudo-differential Operators, Princeton Univ. Press, Princeton, NJ, 1981.

[Tr] F. TREVES, "Introduction to Pseudo-differential and Fourier Integral Operators," Vol. 1, Plenum, New York, 1980.

[Wi] H. WIDOM, "Asymptotics of Compressions to Spectral Subspaces of the Laplacian," A.M.S. Proc. of Symposia on Pure Math., Vol. XXXVI, 1980.

[Z1] S. ZELDITCH, Pseudo-differential analysis on hyperbolic surfaces, *J. Funct. Anal.*, in press.

[Z2] S. ZELDITCH, Uniform distribution of eigenfunctions on compact hyperbolic surfaces, *Duke Math. J.*, in press.